Exploring Medicaid Data for Research: Potentials and Pitfalls

For CERCIT Workshop
(7-22-2011)
By Xianglin L. Du, Ph.D.
UT-School of Public Health
UThrive in Houston

Medicaid Data – brief review

• See previous helpful slides/notes on Medicaid data overview by TCR team (Melanie William & Cheryl Bowcock) on 3/25/11 seminar.

• Medicaid is health insurance that helps many people who can’t afford medical care pay for some or all of their medical bills.

• Medicaid is available only to people with limited income.

“Good health is important to everyone. If you can’t afford to pay for medical care right now, Medicaid can make it possible for you to get the care that you need so that you can get healthy and stay healthy.”

Medicaid Eligibility

Mandatory Eligibility Groups (copied from https://www.cms.gov)

• Limited income families with children, as described in Section 1931 of the Social Security Act, who meet certain of the eligibility requirements in the states’ Aid to Families with Dependent Children (AFDC) plan in effect on July 16, 1996.

• Medicaid Analytic eXtract (MAX) Data Files and Contents

• Medicaid Data quality and validation studies - reviews

• Potentials/Advantages of Medicaid Data

• Pitfalls/Limitations of Medicaid Data

Optional Eligibility Groups (copied from https://www.cms.gov)

• Infants up to age one and pregnant women not covered under the mandatory rules whose family income is below 185% of the Federal poverty level (the percentage to be set by each state);

• Certain people who are otherwise not eligible for the AFDC, but who otherwise are not eligible for the AFDC, but who otherwise are not eligible for the AFDC, but who otherwise are not eligible for the AFDC;

• Tuberculosis-infected persons who would be financially eligible for Medicaid at the Supplemental Security Income (SSI) level (only for TB-related ambulatory services and TB drugs); and

• Low-income, uninsured women screened and diagnosed through a Center's for Disease Control (CDC) Breast and Cervical Cancer Early Detection Program (NBCCEDP) and determined to be in need of treatment for breast or cervical cancer. (Additional information may be found under Downloads and Related Links Inside CMS listed below).
MAX Data Purpose

- Based on maxintro in http://www.ncvhs.hhs.gov/081119p8.pdf by David Baugh, 11/19/2008
- Purpose
  - Produce data to support research and policy analysis on Medicaid and SCHIP (State Children's Health Insurance Program) populations.
  - MAX is needed because source data are not organized to support research.
  - MAX is derived from the Medicaid Statistical Information System (MSIS) for all states & D.C.
- Major Differences between MAX and MSIS:
  For Medicaid services, MAX combines initial claims, voids and other adjustments to create a “final action event”

Whose data is available in Medicaid?

- The CMS Medicaid enrollment and claims data contains information for Medicaid eligibles who actually do enroll in their state's Medicaid program. CMS Medicaid files cannot, therefore, be used to study individuals who are eligible but not enrolled.
- Q: what is the study population?
  - Eligible vs. Enrolled vs. Enrolled+Used Services

Medicaid MAX Data Files

- Personal Summary File (one record per case):
- Claims (Services) Files (multiple records per case):
  - Inpatient Hospital (IP)
  - Long Term Care (LT)
  - Other Services (OT)
  - Prescription Drug (RX)

MAX Data Files (cont’d)

- What types of data elements are in each of 5 files?
  - Personal Summary File:
    - one record for every individual enrolled for at least one day during the year. The file contains demographic data (e.g. date of birth, gender, race), basis of eligibility, maintenance assistance status, monthly enrollment status, and a utilization summary.
  - Inpatient File:
    - complete stay records for enrollees who used inpatient services including 10 diagnoses, 7 procedures, discharge status, length of stay, and payment amount.
  - Long Term Care File:
    - claims for long term care services provided by Skilled Nursing Facilities (SNFs), Intermediate Care Facilities (ICFs), and independent psychiatric facilities. Fields include facility type, 5 diagnoses, dates of service, and discharge status.
  - Drug File:
    - final action paid drug which beginning with 1996 data required an NDC.
  - Other Therapy File:
    - claims records for all non-institutional Medicaid services, including physician services, lab/X-ray, clinic services and premium payments. As appropriate the claims include 2 diagnoses, 1 procedure, and date of service.

Medicaid Data Quality/Validation - review

- Data quality and validation studies?
  - 5 states (California, Florida, New York, Ohio, Pennsylvania)
  - California TR
  - New York study
  - Washington
  - Michigan
  - North Carolina
  - Texas
  - Many other states with non-tumor conditions Medicaid data validation.

Quality of Medicaid Data

- in 5 States: CA, FL, NY, OH, PA
- OBJECTIVES: To perform exploratory analyses of Medicaid and Medicare data obtained through CMS.
- RESEARCH DESIGN: Obtained data from 1999 to 2000 for 5 large Medicaid programs (California, Florida, New York, Ohio, and Pennsylvania), together with corresponding Medicare data for dual eligible.
- RESULTS: by 1) The number of prescription claims per month –stable or not; 2) hospitalizations rates ni Medicaid vs. supplemental Medicare claims; 3). Diagnostic mis-coding in Medicaid (diagnoses of complications of pregnancy and childbirth were uncommon in men and in women older than 60 years, and prostate cancer diagnoses were uncommon in women).
Quality of Medicaid Data (cont’d)  
(in 5 States: CA, FL, NY, OH, PA)  
  – RESULTS: The number of prescription claims per month was very stable. Because of co-coverage by Medicare, Medicaid data seemed to miss a substantial number of hospitalizations in those age "≥65 years as well as 45-64 years".  
  – The diagnostic miscoding was infrequent: diagnoses of complications of pregnancy and childbirth were uncommon in men and in women older than 60 years, and prostate cancer diagnoses were uncommon in women.”  
  – CONCLUSIONS: In contrast to Medicaid data previously obtained from a commercial vendor, we found no evidence that prescription Medicaid data from CMS were incomplete. Researchers using Medicaid data to study hospital outcomes should obtain supplemental Medicare data on dual eligibles for studies of persons aged 45 years and older.

Medicaid vs. Tumor Registry  
(in California)  
  – Medicaid claims corroborated diagnoses for 73% of breast and 68% of colorectal cancers in CCR.  
  – Medicaid claims confirmed surgery for 67% of CCR’s breast cancers.

### Table 1
Number of 1998 Incident Primary Cancers, By Site Reported to the California Cancer Registry (CCR) and the California Medicaid Program

<table>
<thead>
<tr>
<th>Tumor Site</th>
<th>All 1998 Cancers</th>
<th>1998 Cancers Age 16-24 in CCR, %</th>
<th>% Claims Age 16-24 on Medicaid For CCR Cancer</th>
<th>% Claims Age 16-24 on Medicaid For All 1998 Cancers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>20,094</td>
<td>11,223</td>
<td>54.5%</td>
<td>1.1%</td>
</tr>
<tr>
<td>Lung</td>
<td>19,034</td>
<td>9,202</td>
<td>49.0%</td>
<td>5.3%</td>
</tr>
<tr>
<td>Colon</td>
<td>11,264</td>
<td>7,176</td>
<td>64.0%</td>
<td>20.0%</td>
</tr>
<tr>
<td>Rectum</td>
<td>10,021</td>
<td>6,825</td>
<td>68.2%</td>
<td>13.3%</td>
</tr>
<tr>
<td>Testis</td>
<td>667</td>
<td>617</td>
<td>92.4%</td>
<td>12.6%</td>
</tr>
<tr>
<td>Uterine</td>
<td>5,907</td>
<td>3,135</td>
<td>52.8%</td>
<td>32.0%</td>
</tr>
<tr>
<td>Bladder</td>
<td>9,302</td>
<td>1,445</td>
<td>26.0%</td>
<td>15.6%</td>
</tr>
<tr>
<td>Hepatoma</td>
<td>1,975</td>
<td>806</td>
<td>41.0%</td>
<td>29.0%</td>
</tr>
<tr>
<td>Brain</td>
<td>2,802</td>
<td>602</td>
<td>21.5%</td>
<td>32.0%</td>
</tr>
<tr>
<td>Others</td>
<td>7,035</td>
<td>1,445</td>
<td>20.5%</td>
<td>13.8%</td>
</tr>
<tr>
<td>Missing</td>
<td>2,303</td>
<td>2,303</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Medicaid vs. Tumor Registry

### Table 4

<table>
<thead>
<tr>
<th>Claim Definition</th>
<th>Tumor Site</th>
<th>Cause of Death</th>
<th>% Claims Tumor Site</th>
<th>% Claims Cause of Death</th>
<th>% Claims Tumor Site and Cause of Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any other breast surgery or all breast cancers in the Medicaid records of women who had breast cancer and received breast surgery recorded in the CCR</td>
<td>247</td>
<td>321</td>
<td>12%</td>
<td>13%</td>
<td>20%</td>
</tr>
<tr>
<td>Amongst subgroups enrolled in a managed care plan*</td>
<td>240</td>
<td>321</td>
<td>12%</td>
<td>13%</td>
<td>20%</td>
</tr>
<tr>
<td>Amongst subgroups not enrolled in a managed care plan*</td>
<td>257</td>
<td>176</td>
<td>18%</td>
<td>21%</td>
<td>22%</td>
</tr>
</tbody>
</table>
| Medicaid claims corroborated diagnosis for 73% of breast and 68% of colorectal cancers in CCR. Medicaid claims confirmed surgery for 67% of CCR's breast cancers. Conclusion: Medicaid claims have moderate sensitivity for identifying cancer diagnoses and surgery. Linked registry-Medicaid data can identify indigent patients and the timing of Medicaid coverage.

Medicaid vs. Tumor Registry

### Table 3
Cancer Diagnoses and Procedures in California Medicaid Claims for Patients with Cancer Diagnoses Recorded in CCR. Patients Were Enrolled in Medicaid for at Least Some Portion of 1998 and Age 16-64 at Diagnosis

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>Number of RR's with Cancer</th>
<th>Number of RR's with Medicaid in 1998</th>
<th>Number of RR's with Both CCR and Medicaid in 1998</th>
<th>Number (%) with both CCR and Medicaid in 1998</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>1,249</td>
<td>739 (59%)</td>
<td>199 (16%)</td>
<td>16%</td>
</tr>
<tr>
<td>Lung</td>
<td>970</td>
<td>761 (78%)</td>
<td>192 (20%)</td>
<td>20%</td>
</tr>
<tr>
<td>Colon</td>
<td>307</td>
<td>218 (71%)</td>
<td>61 (20%)</td>
<td>20%</td>
</tr>
<tr>
<td>Rectum</td>
<td>262</td>
<td>175 (66%)</td>
<td>63 (24%)</td>
<td>24%</td>
</tr>
<tr>
<td>Testis</td>
<td>74</td>
<td>50 (67%)</td>
<td>16 (22%)</td>
<td>22%</td>
</tr>
<tr>
<td>Uterine</td>
<td>176</td>
<td>121 (69%)</td>
<td>41 (24%)</td>
<td>24%</td>
</tr>
<tr>
<td>Bladder</td>
<td>158</td>
<td>90 (56%)</td>
<td>29 (19%)</td>
<td>19%</td>
</tr>
<tr>
<td>Liver</td>
<td>168</td>
<td>140 (83%)</td>
<td>41 (24%)</td>
<td>24%</td>
</tr>
<tr>
<td>Colon</td>
<td>100</td>
<td>77 (77%)</td>
<td>25 (25%)</td>
<td>25%</td>
</tr>
</tbody>
</table>

**Notes:**

- Medicaid claims corroborated diagnoses for 73% of breast and 68% of colorectal cancers in CCR.
- Medicaid claims confirmed surgery for 67% of CCR’s breast cancers.
- Conclusion: Medicaid claims have moderate sensitivity for identifying cancer diagnoses and surgery. Linked registry-Medicaid data can identify indigent patients and the timing of Medicaid coverage.
Table 3. Cancer Stage Distribution by Medicaid Enrollment Status,* New York State, 2002–2006

<table>
<thead>
<tr>
<th>Stage†</th>
<th>Lung</th>
<th>Colorectal</th>
<th>Female</th>
<th>Breast</th>
</tr>
</thead>
<tbody>
<tr>
<td>In situ</td>
<td>—</td>
<td>7.3</td>
<td>8.8</td>
<td>15.4</td>
</tr>
<tr>
<td>Local</td>
<td>15.3</td>
<td>18.7</td>
<td>27.7</td>
<td>32.8</td>
</tr>
<tr>
<td>Regional</td>
<td>22.0</td>
<td>22.7</td>
<td>35.4</td>
<td>33.4</td>
</tr>
<tr>
<td>Distant</td>
<td>45.3</td>
<td>42.8</td>
<td>16.5</td>
<td>14.8</td>
</tr>
<tr>
<td>Unknown</td>
<td>17.3</td>
<td>15.7</td>
<td>13.1</td>
<td>10.2</td>
</tr>
</tbody>
</table>

* Enrolled in Medicaid at any time during the 2001–2008 period.
†† SEER Summary Stage 2000.

Medicaid vs. Tumor Registry (in New York) (cont’d)


– Conclusions. The linked dataset will permit detailed analysis of cancer treatment and cancer treatment disparities among historically understudied groups. The linkage has also resulted in improvements in Cancer Registry quality through the identification of errors and missing values.

Medicaid-Medicare Data (in Michigan)


– METHODS: Medicaid and Medicare administrative data were merged with the Michigan Tumor Registry to extract a sample of patients who had resection for a first primary colon tumor diagnosed between January 1, 1997, and December 31, 2000 (n = 4765).

– RESULTS: Relative to Medicare patients, Medicaid patients were less likely to initiate chemotherapy (odds ratio, 0.50; 95% confidence interval, 0.39-0.65) or complete chemotherapy.

Medicaid Data in North Carolina


<table>
<thead>
<tr>
<th>Comorbidity</th>
<th>&lt;65</th>
<th>65-74</th>
<th>75+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes (1+)</td>
<td>89 (62.9%)</td>
<td>99 (55.6%)</td>
<td>209 (18.7%)</td>
</tr>
<tr>
<td>No (0)</td>
<td>117 (75.2%)</td>
<td>78 (47.4%)</td>
<td>100 (18.0%)</td>
</tr>
</tbody>
</table>

Medicaid Data Validation – high disenrollment after Dx


• METHODS:
  • Medicaid enrollment and claims files were linked to the Washington State Cancer Registry to identify all Medicaid enrollees with breast, cervical, lung, colorectal, and prostate cancer between 1997 and 2002.
  • RESULTS:
    • We identified 5009 newly diagnosed cancer patients covered by Medicaid, approximately 13% of the total cases diagnosed in subjects less than 65 years of age in Washington State. The majority, 2866 (57%), enrolled in Medicaid around the time of diagnosis; the remainder had been enrolled at least 3 months before diagnosis. Persons enrolled at diagnosis had later-stage cancer; those enrolled before diagnosis had more noncancer comorbidities. Overall, 18% had disenrolled by 6 months after diagnosis; 34% by 1 year; and 54% by 2 years.

• CONCLUSIONS:
  • Medicaid patients with cancer in Washington State experience a high rate of disenrollment within 1 year after diagnosis. Further research is needed to determine whether disenrollment compromises initial therapy or follow-up care.
Medicaid Data Validation – high rate of incomplete treatment

- Aims: "We evaluate radiation treatment completion rates for Medicaid enrollees aged 18-64 diagnosed with breast cancer, using data from the Washington State Cancer Registry linked to Medicaid enrollment and claims records, we identified Medicaid enrollees diagnosed with breast cancer from 1997 to 2003 who received BCS. Among the 402 women who met inclusion criteria, 105 (26%) did not receive any radiation."
- Results:....
- Among those who received at least one radiation treatment, 65 (22%) failed to complete therapy and 71 (24%) patients had at least one 5 to 30 day gap in treatment.
- Conclusions:....
- A substantial proportion of Medicaid-insured women who are eligible for high rate of incomplete treatment.

Medicaid Data Validation (in Texas)


Table 1. Treatment Data in Women with Cervical Cancer in 1999-2001: Texas Cancer Registry Vs. Medicaid Claims Data

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Majority Level</th>
<th>Non-Hispanic Whitea</th>
<th>Non-Hispanic Black</th>
<th>Hispanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment: Medicaid††</td>
<td>N=360</td>
<td>N = 157</td>
<td>N= 340</td>
<td></td>
</tr>
<tr>
<td>No treatment (REF)</td>
<td>158 (38.76)</td>
<td>89 (45.70)</td>
<td>129 (85.43) (p=0.34)</td>
<td>202 (85.88) (p=0.17)</td>
</tr>
<tr>
<td>Any treatment</td>
<td>158 (41.24)</td>
<td>79 (42.90) (p=0.02)</td>
<td>100 (64.12) (p=0.00)</td>
<td></td>
</tr>
<tr>
<td>Type of Treatment: Medicaid††</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No: Treatment (REF)</td>
<td>158 (38.76)</td>
<td>89 (45.70)</td>
<td>129 (85.43) (p=0.34)</td>
<td>202 (85.88) (p=0.17)</td>
</tr>
<tr>
<td>Any: Treatment</td>
<td>158 (41.24)</td>
<td>79 (42.90) (p=0.02)</td>
<td>100 (64.12) (p=0.00)</td>
<td></td>
</tr>
<tr>
<td>Type of Treatment: TCR††</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No: Treatment (REF)</td>
<td>158 (38.76)</td>
<td>89 (45.70)</td>
<td>129 (85.43) (p=0.34)</td>
<td>202 (85.88) (p=0.17)</td>
</tr>
<tr>
<td>Any: Treatment</td>
<td>158 (41.24)</td>
<td>79 (42.90) (p=0.02)</td>
<td>100 (64.12) (p=0.00)</td>
<td></td>
</tr>
</tbody>
</table>

- this was similar pattern/finding to that of SEER and/or Medicare data.

Potentials/Advantages of Medicaid Data

- Large # of community-based cases for low-income populations by states and across the US states.
  e.g. 'in any given month there are > 3 million clients enrolled in the Texas Medicaid Program' (source: Bowcock@TCR);
- Study <65 as well as >=65 yrs of low-income populations (children, pregnant women, elderly, & people with disability);
- Help identify additional CA cases (that may be missed by tumor registry);
  e.g., (source: Bowcock@TCR): 45,226 Medicaid records contained a malignant cancer-related ICD-9 diagnosis code, but there was no record in TCR data.
  - 4,086 prostate (9.0% of the non-linked Medicaid records)
  - 5,853 female breast (12.9%)
  - 3,211 colorectal (7.1%)
  - 756 melanoma (1.7%)

Potentials/Advantages of Medicaid Data (2)

- Enhance the treatment data:
  - that were not recorded in tumor registry;
  - e.g., radiation therapy for breast ca (source: Bowcock@TCR):
    Reported to TCR. 2,361; and not reported 1,183 (33%).

Potentials/Advantages of Medicaid Data (3)

- Provide other important treatment information (not covered by other programs), e.g. hormone therapy for breast ca (not covered by Medicare, before Part D in 2006),
Potentials/Advantages of Medicaid Data (4)
- Valuable to evaluate the CDC’s National Breast and Cervical Cancer Early Detection Program (NBCCEDP): uninsured and underinsured women at or below 250% of federal poverty level; ages 18–64 for cervical screening; ages 40–64 for breast screening;
- CONCLUSION: Treatment patterns among Georgia Medicaid cases appear appropriate to stage but 18% with invasive cervical cancer received no cancer treatment, although Medicaid enrolled.
- Texas Breast Cancer Screening, Early Detection and Treatment Program Outcomes Study (PI: Dr. Begley at SPH) funded by CPRIT, to understand the benefits and failings in the Texas Breast and Cervical Cancer Services (BCCS) program and develop innovative strategies for program improvement and expansion to a larger proportion of the low income population.

Pitfalls of Medicaid Data (cont’d-4)
- Other limitations:
  - No reliable information on defining cancer stage;
  - Incomplete data for Medicaid enrollees who belonged to HMOs, other insurance such as employer’s health plan and those who received care in Veterans Affairs hospitals.
Q: Can managed care services be studied with the CMS Medicaid claims data files? (from CMS-ResDAC website):
  - There are 2 approaches to identifying Medicaid beneficiaries who were covered under a managed care plan.
    - The Personal Summary File indicates if a beneficiary was in a managed care plan with coverage purchased by the state, and this information is indicated for each month.
    - It may be necessary for other types of plans to determine if the managed care is medical, dental, transportation or some other form. This information can be obtained by examining claims in the Other Therapeutic Claims file. Claims with a Type of Service = ‘20’ (premium payment) indicate beneficiaries covered by a managed care plan.
Thank you for Attention!

Questions/Comments